ELSEVIER

Genomics and proteomics
From genes to function: advances in applications of chemical and systems biology

Editorial overview

Matthew Bogyo and Benjamin F Cravatt

Current Opinion in Chemical Biology 2007, **11**:1–3

Available online 10th January 2007

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2006.12.029

Matthew Bogyo

Department of Pathology and Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305-5324, USA
e-mail: mbogyo@stanford.edu

Matthew Bogyo received his BSc degree in chemistry from Bates College in 1993 and a PhD in biochemistry from the Massachusetts Institute of Technology in 1997. He became a Faculty Fellow at the University of California, San Francisco in 1998. In 2001, he moved to Celera Genomics to head the Department of Chemical Proteomics until 2003 when he made the transition back to academics as an Assistant Professor in the Department of Pathology at Stanford University. His research interests focus on the use of small molecules to study the functional roles of proteases in human diseases, with an emphasis on cancer and malaria.

Benjamin F Cravatt

The Skaggs Institute for Chemical Biology and Departments of Cell Biology and Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA
e-mail: cravatt@scripps.edu

Benjamin Cravatt obtained his undergraduate education at Stanford University, receiving a BS in the Biological Sciences and a BA in History. He then received a PhD in Macromolecular and Cellular Structure and Chemistry from The Scripps Research Institute (TSRI) in 1996. Cravatt joined the faculty at TSRI in 1997 and is currently a Professor in the Skaggs Institute for Chemical Biology and the Departments of Cell Biology and Chemistry. His research group is interested in mapping biochemical pathways in human disease using advanced proteomic and metabolomic technologies.

In the so-called post-genomic era we face a formidable challenge as scientists – to make sense of the huge amount of sequence information generated by an ever-growing list of completed genome projects. Although it was once thought that sequencing of the human genome would immediately lead to advances in our understanding of human health and disease, this goal remains largely unrealized. Even with significant progress in the annotation of genes based on predicted function, the overwhelming complexity of gene regulation and protein expression has made it difficult to globally map regulatory pathways that control basic cellular processes. Confronted with these hurdles, the past ten years have seen a rapid increase in the development of new technologies that enable systems-wide analysis of cellular processes. These technological advances depend on the tools of many disciplines, including computational biology, chemistry, protein biochemistry and mass spectrometry. It has become clear that we must make use of strengths in each of these fields to begin to reap the rewards of the genome sequencing efforts. In this volume, we have chosen a number of topics that highlight advances in technologies that enable scientists to approach biological questions using global rather than single gene and/or protein methods.

To begin this volume, **Thomas and co-workers** provide a comprehensive look at methods to assign function to genes in sequenced genomes. This review is an ideal starting place for the volume as it outlines computational efforts that make use of raw sequence information, curation of published literature, inheritance prediction and pathway mapping to better define gene function. Although the biological function of genes is something that can, in principle, be gleaned from basic homology of sequences to other related proteins, this link is often tenuous at best. Specific functional annotations can be significantly strengthened by analysis of homology among multiple species, something that is now possible with the completion of many new genome sequencing projects. Even more importantly, we must rely on the published literature to be able to make annotations that can be backed up with real experimental data. Finally, we need to develop methods to link this information into networks that are defined by the biological process in which they are involved. The addition of new search algorithms and network databases is already beginning to enable the transition from gene sequences to comprehensive maps of complex biological processes.

Transitioning from 'sequence gazing' and annotation of predicted function at the level of gene sequence homology, we have selected a few reviews that shift the focus to proteins. One way to enhance our understanding of protein

function is to develop technologies that enable direct imaging of protein expression and localization. The advent of a host of stable fluorescent reporter molecules has provided the cell biologist with veritable toolbox of reagents that enable specific proteins to be tagged and monitored visually. Although simple fusion protein techniques have opened the door to many new cell biological studies, it still remains difficult to dynamically image protein trafficking and localization for many proteins. **Foley *et al.*** outline some of the most recent advances in site-specific modification of proteins with reporters. This review highlights the diversity of methods to selectively tag proteins, thereby facilitating their analysis by imaging, biochemical and proteomic methods. The future of this field depends heavily on creative chemists and biochemists to continue to develop new techniques that will enable this approach to be adapted to system-wide studies of protein function.

Continuing with the topic of imaging using reporter tags, we have selected two reviews that focus on analytical methods to image protein and small molecules in complex proteomic samples. In the first review, **Sadaghiani *et al.*** outlines recent advances in protein labeling methods using small-molecule activity-based probes. Similar to the tools described by Foley *et al.*, this approach enables proteins to be studied using reporters that include epitope tags, affinity tags and fluorescent reporters. The difference in this method is that the labeling of the protein is facilitated by a small-molecule probe that binds to the target protein using an activity-dependent chemical reaction. Thus, probe modification provides not only a way to visualize a given protein but also an indication of its enzyme activity. Because many enzymes are regulated by a complex set of post-translational controls, the use of activity-based probes greatly enhances our ability to dissect protein function by focusing analysis on changes at the level of enzyme activity.

Although activity-based probes provide a means to monitor sets of functionally related proteins, these reagents can only be applied to enzymes for which specific probes can be designed. The review by **Reyzer and Caprioli** outlines mass spectrometry methods that enable proteins and small molecules to be imaged in tissues and whole organisms using a direct mass spectrometry-based readout. Using this technique it becomes possible to perform non-biased profiling of both proteins and small-molecule metabolites in whole tissue sections. This approach has the potential to facilitate the discovery of biomarkers, high-resolution mapping of protein expression *in vivo* and evaluation of small-molecule pharmacodynamic properties.

In addition to finding ways to globally monitor protein expression, localization and enzyme activity, other systems-based approaches have dealt with the problem of complexity by focusing on specific sub-sets of the genome. The reviews by **Schilling and Overall**, and **Diamond** highlight some of the recent advances in the study of families of enzymes, known as proteases, that regulate cellular function by degrading other proteins. Schilling and Overall outline several exciting new mass spectrometry-based methods that enable global monitoring of protease expression, activation and substrate profiling. These methods are likely to have a dramatic impact on our understanding of how groups of related enzymes (i.e. proteases) function in regulatory networks. Building on this theme, Scott Diamond highlights some of the advances that can be used to profile the substrate specificity of an expressed protease. Using techniques ranging from screening of libraries of diverse fluorogenic substrates to profiling cleavage specificity using phage display and other genetic tricks, it is now possible to assign specificities to proteases that can then be used to search for substrates. These technical advances are already being translated to other classes of enzymes and will probably have an important role in helping to establish functional networks in the genome.

Continuing along the lines of focused, systems-wide approaches, two reviews address advances in the study of proteins that are post-translationally modified with glycans. In the first review, **Bond and Kohler** highlight new methods for the isolation and biochemical evaluation of glycoproteins. These methods make use of a number of chemical tricks in which sugar monomers are modified with chemical 'handles' to enable the isolation of glycoproteins as well as the mapping of glycosylation sites. These methods rely heavily on chemical synthesis methods as well as on analytical methods such as mass spectrometry. Such systems-wide analysis of glycoproteins will undoubtedly lead to a better overall understanding of the roles of glycosylation in diverse cellular processes.

In the second review focused on glycoproteins, **Timmer *et al.*** discuss recent advances in the application of probes to study the enzymes that regulate glycosylation. Relating back to the review by Sadaghiani *et al.* that introduces small-molecule activity-based probes, this review highlights some valuable additions to the list of probes that can be used to profile families of related enzymes in the genome. In this case the focus is on proteins that have important roles in the post-translational modification of glycosylation, a process that has crucial functional role in both normal cellular physiology and disease pathology.

Some of the most exciting classes of proteins also represent the most technically challenging targets for large-scale profiling. Post-translationally modified proteins are a group that both fascinate and frustrate biochemists because of their key biological functions and daunting molecular complexity. Histones and their

myriad modifications, including acetylation, methylation, phosphorylation and the many combinations thereof, represent a prototype case study of the remarkable manner in which cells convert an individual gene sequence into hundreds of functionally distinct protein species. **Hunt and co-workers** describe exciting breakthroughs in mass spectrometry methods that are enabling, for the first time, large portions of sequence to be read from intact histone proteins. These efforts promise to deliver the first method by which combinations of post-translational modifications can be profiled on single proteins.

Complementing the aforementioned genomic and proteomic methods, investigators are also employing small molecules as probes for the global interrogation of protein function. The discovery of specific pharmacological tools to perturb protein function is a well-recognized challenge and one that is creatively addressed by the small-molecule microarray (SMM) technology described by **Duffner *et al.*** These SMMs have the potential to offer the first general platform for the systematic discovery of protein–ligand pairs on a genome-wide scale. Bioactive small molecules can also be unearthed using cell-based screening, a topic that is often referred to as chemical genetics and is covered by two captivating reviews in this series. **Gangadhar and Stockwell** describe the identification of small molecules that regulate cell-death pathways in mammalian systems through novel mechanisms. These previously obsure cell-death pathways uncovered by chemical genetics experiments might offer new therapeutic targets for the treatment of diseases such as cancer. Chemical genetic studies have also illuminated many exciting aspects of plant biology, a subject that is reviewed by **Kaschani and van der Hoorn**. As the authors describe, plants possess special metabolic, cell, and systems biology pathways that have been found to be sensitive to small-molecule perturbation. Finally, natural small-molecule signaling pathways, such as those mediated by nuclear hormones, have also proven to be a rich source of discovery and innovation for chemical biologists, as delineated by **Biggins and Koh**. Nuclear hormone systems not only are a rich source of drug targets, but also have provided fertile ground for researchers interested in engineering innovative tools to control gene expression with unparalleled versatility and selectivity.

Taken together, this collection of reviews, although diverse in the overall subject matter and technological focus, is unified by the overarching goal of dissecting biological function on a global scale. Scientists have learned a great deal since the completion of the human genome project less than a decade ago. Perhaps most importantly we have learned that it is going to take creative input from many disciplines to begin to 'translate' the genome and unlock the great potential this information has for the betterment of mankind. We hope that you will agree that this compilation of reviews provides an up-to-date survey of some of the most exciting technological advances in systems biology that are likely to help us reach this goal.

Functional genomics attempts to understand this dynamic variability in which genes are expressed. This can lead into epigenetics, transcriptomics, proteomics, microRNA interactions, and more. Epigenetics can tell us about which sections of the genome have been activated or suppressed through the histone code or methylation.Â Genomics, transcriptomics, and proteomics are all incredibly important in understanding how biological systems work. 3. Related Answer.Â This shows that results drawn from Transcriptomics and Proteomics do not answer the same question. Abstract: Advances in genomics and proteomics increasingly contribute to the understanding of signal transduction pathways that control growth, differentiation, and death of cells. Since defects in these processes may result in the expression of inherited and or acquired disease, the identification of candidate disease genes and modifier genes by parallel use of genotyping together with an integrated study of gene expression and metabolite levels is instrumental for future health care.Â Gene interaction networks have recently been demonstrated, in which hub genes, that is, genes that show the highest level of interactions with other genes, play a special role. Put more simply, proteomics analyzes the structure and function of biological systems. For example, the protein content of a cancerous cell is often different from that of a healthy cell. Certain proteins in the cancerous cell may not be present in the healthy cell, making these unique proteins good targets for anti-cancer drugs.Â The study of the function of proteomes is called proteomics. A proteome is the entire set of proteins produced by a cell type.Â Proteomics complements genomics and is useful when scientists want to test their hypotheses that were based on genes. Even though all cells of a multicellular organism have the same set of genes, the set of proteins produced in different tissues is different and dependent on gene expression. Chemical Biology and Proteomics. Developmental and Stem Cell Biology. Genetics, Genomics and Gene Regulation. Immunology and Infectious Diseases. Neurobiology.Â We use 2D and 3D model systems for in vitro investigations. We have also generated novel transgenic mice for metastasis studies in vivo. Our goal is to prevent any future deaths due to breast cancer. Rachel Green. Work in the Green lab is centered on the ribosome, and we are interested in deciphering the molecular mechanisms that are at the heart of protein synthesis and its regulation across biology. This focus allows us to still think about the earliest evolutionary steps that led to life on earth, but in a system where biological questions drive the experiments.